



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Reshaping quality and equity

**Citation for published version:**

Aaron Benavot & Smith, W 2020, Reshaping quality and equity: Global learning metrics as a ready-made solution to a manufactured crisis. in A Wulff (ed.), *Grading Goal Four: Tensions, Threats and Opportunities in the Sustainable Development Goal on Quality Education*. Brill, pp. 238-261.  
<<https://brill.com/view/book/edcoll/9789004430365/BP000009.xml>>

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Grading Goal Four

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Reshaping Quality and Equity: Global Learning Metrics as a Ready-Made Solution to a Manufactured Crisis

*Aaron Benavot and William C. Smith*

## 1 Introduction

Ranking countries on a global scale of learning has become a top priority in the education world. The introduction of global learning metrics (GLM) effectively transforms ‘conventional’ discussions of education progress, which have focussed on enrolment or completion rates, gender parity, and out-of-school children. While there is no common definition of a GLM, it typically refers to a single global scale in which measures of learning from different standardised assessments are placed (Hanushek & Edwards, 2017). The UNESCO Institute for Statistics (UIS) discusses, in theory, an ideal GLM based on a perfectly equated learning assessment programme (UIS, 2018g). It is unlikely that such an ideal GLM will ever come into existence. Meanwhile, UIS is developing guiding tools and definitions to support the alignment and comparability of results from different assessments in relevant domains and at different education levels. This would mean that nationally representative assessment programmes would begin to use shared definitions and linking methodologies to create a common format of reporting (a global scale or metric) in a transparent way (UIS, 2018g). The Australian Council for Educational Research (2019) refers to a GLM as a ‘universal learning progression’ in which student achievement on any national, regional, or international learning assessment can be converted into universal learning progression units.

The fascination with GLMs shifts the focus to the outcomes of schooling and embraces the mantra of results-oriented policymaking. GLMs also enable countries to report progress on the Sustainable Development Goal on Education (SDG 4), specifically the first target (4.1), which calls on countries to ‘ensure that all girls and boys complete free, equitable and quality primary and secondary education *leading to relevant and effective learning outcomes*’ (WEF, 2015, p. 20, emphasis added).

The effort to compare national learning outcomes on a universal scale has been spearheaded by UIS in its official capacity to collect cross-nationally

comparable data to measure SDG 4 targets. Supported by major bilateral and multilateral donors, UIS has been working overtime to construct and report global, regional, and national estimates of the percentages of children/young people who achieve a minimum proficiency level in reading and mathematics in primary and secondary education (the global indicator of Target 4.1). For example, UIS currently reports reading proficiency data for students at the end of lower secondary education (typically grade 8 or 9) for almost 90 countries.<sup>1</sup> Such global coverage of learning levels would have been unthinkable two decades ago.

There is no consensus on the technical procedures to combine information from different assessment platforms.<sup>2</sup> Nevertheless, the overall message is crystal clear: first, all countries in the world should conduct nationally representative learning assessments of children and youth, preferably by participating in international assessments, in order to determine learner proficiency levels in reading and mathematics; and second, it is desirable to combine select results from such assessments and map them onto a global learning scale. In effect, learning should be seen as independent of national context – for example, independent of education structure, curricular policy, language of instruction, and level of development.

This chapter critically interrogates on-going efforts to establish and legitimate global learning metrics. It highlights how and why the massive push to ensure that all students worldwide demonstrate measurable proficiencies in reading and mathematics has emerged, and with what consequences for the broader SDG 4 agenda, especially equity issues. Drawing on the growing rhetoric of a ‘global learning crisis’ and informed by innovative yet problematic technical work, we argue that the powerful movement to construct GLMS has several ‘unintended’ outcomes. These include the effective narrowing of the comprehensive global agenda on education (SDG 4), the undermining of a carefully negotiated country-led process to promote lifelong education opportunities for all, the devaluing of learning that is not measurable or comparable, and the weakening of the principle of educational equity.

## 2 The Rise of Large-scale Comparative Assessment and the Quantification of Education Outcomes

Since the 1960s standardised learning assessments have seen a dramatic upswing in usage. In 1961, the International Association for the Assessment of Educational Achievement (IEA) completed its first pilot study, concluding that cross-nationally comparable results were possible (Pettersson, Popkewitz,

& Lindblad, 2016). Momentum for comparison grew in the 1970s and 1980s as more researchers believed education systems could be systematically compared with each other (Kamens, 2013). Between 1960 and 1989, 43 international surveys of academic achievement were conducted (Heyneman & Lee, 2014). During the 1990s and 2000s, participation in regional and national assessments increased rapidly (Kamens & Benavot, 2011). Fuelling the motivation for standardised testing was an assumption that 'the quality of educational practices can be unambiguously quantitatively measured and that such measures are sufficiently precise and robust to be aggregated into policy-relevant rankings' (Meyer, 2017, p. 17).

The establishment of the OECD's Programme for International Student Assessment (PISA) in 1999 provided a robust platform for the comparison of student learning. Between 1999 and 2012, participation in PISA and other international assessments increased by 50% (Smith, 2014). Overall, by 2008 nearly three-quarters of developing countries had participated in at least one national, regional, or international assessment (Kamens & Benavot, 2011).

Beyond the unprecedented increase in the number of tests conducted, there has been a shift in the intention and ownership of such assessments. Pizmony-Levy (2013) highlights the relative decline in the number of researchers participating in the IEA's General Assembly, replaced by individuals affiliated with, or officially representing, national governments. Some have also pointed to the changing purposes of testing, with greater emphasis on using outcomes for accountability (Smith, 2014). The pattern of increased government involvement made clear that national education policymakers viewed the assessment of learning as 'an important, perhaps a key, strategy for improving educational quality' (Chapman & Snyder, 2000, p. 457).

The rise of learning assessments mirrored an increased reliance on quantitative measurement rather than qualitative judgement. Buttressed by a belief in meritocracy and positivism, and an imperative to avoid subjective value judgements and perceptions of discrimination, some trusted that 'the only hierarchy that can be accepted is based on meritocratic ideas aggregated from evaluations of the performance of individuals' (Pettersson et al., 2016, p. 180). Positivism suggests that true levels of merit can be objectively measured (Abraham, 1994). Numbers are seen as 'technical, objective, and calculable and embodying the idea of giving all equal chances and representation' (Pettersson et al., 2016, p. 184) with comparable data replacing personal judgement (Muller, 2018). The efficiency movement in the early 20th century brought positivism into education, advocating for a scientisation of education with standardised and quantified best practices replacing teachers' intuition (Meyer, 2017). In the past thirty years, economic globalisation has pressured countries to assess

the competitiveness of their education systems and labour forces (Kamens & Benavot, 2011). More broadly, formal institutions are ‘increasingly ... subjected to performance measurements that define success or failure according to narrow and arbitrary metrics’ (Muller, 2018).

Increased country participation in learning assessments reflects a global environment in which education policies are increasingly diffused, borrowed, and contextualised (Steiner-Khamisi, 2004). Indeed, countries more integrated into world society are more likely to test students (Kamens & McNeely, 2010). In addition, in what some describe as the ‘global education compact’ (Daun & Mundy, 2011; Mundy, 2006), formerly ideologically opposed institutions, such as the World Bank and UNESCO, are working together toward a merging of agendas. This convergence can be seen in the *Education 2030 Framework for Action* (WEF, 2015), where certain guidelines such as fair and inclusive education are more aligned with humanist approaches supported by UNESCO, while others, such as defining education quality through testing, derive from an instrumental or neoliberal paradigm commonplace in the World Bank (Akkari, 2018). Sahlberg (2011) refers to a ‘global education reform movement’ that reinforces neoliberal principles and reforms such as increased decentralisation, standardisation, and privatisation. Learning assessments, drawing on the ‘global testing culture’, derive from and further encourage such education reforms (Smith, 2016a). This testing culture draws scripts and models of expected behaviour for all education stakeholders, which shape how education is understood and valued. It thus becomes common sense that ‘testing is synonymous with accountability, which is synonymous with education quality’ (Smith, 2016b, p. 7).

### 3 Debating the Post-2015 Agenda for Education

As discussions over post-2015 priorities were held, two overarching camps – with different foci and underlying ideologies – sought to influence the direction of the global education goal and targets (see Chapter 9 by Yusuf Sayed and Kate Moriarty). The humanistic camp pushed strongly for education that was fee-free and inclusive (Unterhalter, 2019). Based on a rights-based approach that placed government as the primary duty bearer, this camp focussed on issues related to equity, social justice, and nondiscrimination (Brissett & Mitter, 2017). By contrast, the economic camp, undergirded by human capital theory, highlighted education’s role in economic development and tied education quality to labour force demands and occupational opportunities. The main purpose of education, according to the camp, is utilitarian: ‘preparing children

to work within an established socio-economic order with the ultimate goal of achieving economic growth' (Brissett & Mitter, 2017, p. 195).

In the consultation process over the emergent global goal on education, debates between the two camps ensued. In 2013, at the Thematic Consultation on Education in Dakar, Senegal, the outcome document advanced a limited view of quality as meeting minimum standards in reading, writing, and counting at the primary level with an overarching emphasis on learning outcomes (UNESCO, 2013d; Unterhalter, 2019). Subsequently, as Unterhalter (2019) found in her review of the lead-up to SDG 4, expert-led consultations tended to emphasise links between inadequate learning and poor economic growth. Wider consultations initiated by the Open Working Group viewed education more comprehensively, emphasising provision at all levels and providing broader definitions of quality that included enabling conditions and diverse learning outcomes, including for sustainable development and global citizenship. The humanistic approach illustrated through the Open Working Group was, in part, fuelled by the active participation of civil society organisations (for more details, see Chapter 2 by Antonia Wulff).

#### 4 Key Concerns of the Two Camps: Equity and Learning

Equity and learning represent core concerns in both camps. Where they differ is how the issues are framed and to what purposes. The utilitarian view of education emphasises a narrower array of school-based learning outcomes, typically foundational skills, measured rigorously and assessed frequently, which serve as the basis of evidence-based reforms. The humanistic camp emphasises equity and rights-based approaches in education and a broader conception of quality, including inputs, processes, and outcomes. Assessing learner experiences and an array of learning outcomes, both inside and outside of formal schooling, as well as the provision of qualified, prepared, and duly compensated teachers, are key to this view.

For the utilitarian camp, education for all had become 'learning for all', an (some would say 'the') overarching policy priority in which the measurement and assessment of learning took centre stage. Assessments, especially those that lent themselves to cross-national comparison, would enable policymakers to identify policies that improve the skills and competencies of current students and enhance future workers' competitiveness in the global economy. The humanistic camp had a more ambivalent attitude toward assessments, since it shifts the focus from enabling conditions and quality teaching to test scores as the privileged criteria for policy formulation. That said, disaggregated

data from learning assessments might be beneficial insofar as they shine a light on the distinct learning challenges facing marginalised and excluded children.

To effectively address the learning challenges faced by the least advantaged populations, assessments would need to collect detailed information about multiple disadvantaged groups. Concerns were voiced as to which groups to include in assessments and which to leave out (Doble, 2015). For example, should assessments be school-based (thereby excluding children not enrolled in school) or household-based (thereby excluding those not living in a household)? Should they go beyond households and sample orphans or those living in institutionalised settings? Should they include 'unregistered' children or those living in 'illegal' refugee or migrant settlements? For equity purposes the sampling frame and sample size of learning assessments are critical issues since they determine the (non)representation of at-risk groups. This is especially true for learners with intersecting disadvantages – for example, girls with disabilities or linguistic minorities who live in rural villages (Lockheed & Wagemaker, 2013). Furthermore, disadvantage is often context-specific, requiring country input and attention to salient groups (Benavot, 2018b).

A longstanding critique of cross-national assessments is their inability to capture meaningful differences among learners who score at the lower end of a learning scale (Lockheed & Wagemaker, 2013). In PISA, for example, two-thirds of countries that scored below the OECD average in 2009 were low- and middle-income countries. In Peru, 82% of students fell below the 400-point mark. Such students were deemed illiterate, which means that the assessment provided little useful policy information (Lockheed & Wagemaker, 2013). Results from the 2015 PISA indicated that the reading score of the typical poor country was below the fifth percentile of OECD countries. This percentile is considered close to 'special needs' (Crouch, 2017). This suggests that the lowest levels in assessments like PISA or Trends in International Mathematics and Science Study (TIMSS) are too high for most students in low-income countries (Winthrop and Anderson Simons, 2013). Given this lack of detailed information at the lower end of the skills spectrum, analyses to identify associated factors may be inaccurate (Lockheed & Wagemaker, 2013). The IEA suggests that accuracy declines when students score less than 30% correct (Crouch, 2017). Lengthening the assessment could provide useful information about students scoring at the bottom, but this may not be a feasible, or complete, solution (Crouch, 2017).

In addition to detailed information on lower learning levels, understanding the determinants of learning among disadvantaged learners requires extensive background information (Klemenčič & Mirazchiyski, 2018). This is usually accomplished through companion surveys completed by students, teachers,



school administrators, and/or parents. Unfortunately, this information gets little attention in public policy discussions, thereby missing the context-specific obstacles facing marginalised groups (Winthrop and Anderson Simons, 2013). Using data for equity purposes also entails that teachers and school leaders have access to data in a format they understand and can use (Rose, 2016). Summative assessments, especially those linked with accountability measures, assume that all students start at the same development level and thus make it difficult to tailor interventions to specific groups of learners (Ahsan & Smith, 2016). Furthermore, it is erroneous to believe, as many decisionmakers do, that policies found to be effective for the 'average' or typical learner will be equally effective in addressing the needs of learners from marginalised groups (Benavot, 2018b). Detailed data and specialised analyses are critical for identifying more or less effective policies for marginalised learners.

#### 4.1 *Manufacturing a 'Crisis' in Learning and the Push for Reform*

The language and narrative employed to frame the results of learning assessments also distinguishes the two camps. Words like 'crisis' and 'shock' have been used in the past to describe the 'appalling' state of affairs in education. In the 1960s and again in the 1980s, Philip Coombs wrote extensively about conditions fostering a 'world educational crisis', especially in the Global South (Coombs, 1968, 1985). In 1983, *A Nation at Risk* garnered extensive media attention in the US, claiming 'a rising tide of mediocrity [in education] that threatens our very future as a nation and a people' (US National Commission on Excellence in Education, 1983, p. 1) and touched off a wave of local, state, and national reforms. In 1993, Alan Rogers referred to the 'world crisis' in adult education, especially in relation to adult literacy (Rogers, 1993). In 2000, the results of the first PISA assessment revealed a lacklustre performance of German students. The public 'shock' then triggered intensive public discussion and scholarly debates about the need for extensive education reform. The 2013 EFA Global Monitoring Report highlighted 'the global learning crisis', estimating that, regardless of whether they have ever attended school, 250 million children could not read, write, or count well and that 775 million adults lacked basic reading and writing skills (UNESCO, 2014c, p. 191). In short, while the crisis hyperbole has a long history in education circles, its prevalence appears to be increasing.

Among those who employ a 'crisis' narrative in reference to learning, the diverse contexts in which learning deficits are created and fostered are often minimised. The purported existence of a 'global learning crisis' leaves little room for nuance. It reduces the issue into a simple dichotomy: some education



systems are successfully producing students who achieve high average test scores on assessments, while most systems are ineffective, failing, or both, and in need of significant restructuring. Such narratives are commonly mobilised by those concerned with economic growth or global competitiveness. They tend to overlook deeper conditions that contribute to low or unequal learning levels and would need to be addressed if real improvements were to be realised.

In many contexts, the media latch onto and amplify the crisis narrative. They flash ominous headlines with country rankings on television and radio. Driven by directives that require the rapid production of simplified, newsworthy material (Yasukawa, Hamilton, and Evans, 2017), media reports on education tend to be overly negative and emphasise quality or excellence over equity (Baroutsis & Lingard, 2017). Results from international large-scale assessments, especially PISA, have garnered growing media attention. Often presented in international league tables or country rankings (see following paragraphs), the stories ‘reduce the complexity of PISA findings to simple messages that are aimed at changing or reinforcing particular perceptions of education and influencing the decisions of policy-makers’ (Sellar, Thompson, & Rutkowski, 2017, p. 25). The OECD and the World Bank have encouraged the use of such ‘catalyst data’ (Sellar et al., 2017, p. 19) to ‘spur action’ (World Bank, 2017b, p. 93). The naming and shaming approach of countries through league tables has been described as PISA shock. Reform efforts following the release of PISA results have been documented in countries like Denmark, Germany, Japan, and Portugal (Rey, 2010; Volante, 2015).

Media presentations of country comparisons not only reinforce the notion that education quality can be reflected in a single test score, but often identify who is to blame. Responsibility for poor quality (i.e., low test scores) is laid at the feet of schools and teachers, whereas little consideration is given to either systemic problems (e.g., insufficient funding, substandard school structures, and inappropriate instructional materials) or uneven policy implementation (Kumashiro, 2012). In Turkey, following back-to-back poor performances on the 2003 and 2006 PISA, the Ministry of National Education overlooked structural and systemic issues and placed blame predominantly on teachers and their inability to implement the new curriculum (Gür, Celik, & Özoğlu, 2012). Research on media representation of teachers in Australia, Bangladesh, Oman, Saudi Arabia, and South Africa reinforced a one-sided portrayal of teachers as lazy, unprofessional, and often engaged in misconduct (Alhamdan et al., 2014). This may be shaped in part by the absence of educator voices as experts in media reports on education (Yasukawa et al., 2017).

## 5 The Role of League Tables in Manufacturing Calls for Reform

Whether the learning crisis is real or manufactured, it has taken on a new complexion in the post-2015 era as league tables have gone global. David Edwards, the General Secretary of Education International, suggests that poor learning outcomes are not a shock for those on the front line; teachers are keenly aware of poor learning levels since they labour in overcrowded classrooms and underresourced schools (Hanushek & Edwards, 2017). What is damaging in recent years is how learning challenges and the learning crisis are presented. Global league tables serve as a mechanism by which assessments enact hierarchical control. Country rankings on international assessments become part of an international struggle for developing (and securing) talent (Volante & Ritzen, 2016). Media outlets assume that rankings contain adequate information to draw conclusions about education quality: rankings are 'often the only evidence used in policy debates on education' (Klemenčič & Mirazchiyski, 2018, p. 1). This is despite the limited and relatively uninformative information provided (Klemenčič & Mirazchiyski, 2018) and the consistent misrepresentation or misinterpretation of assessment results (Sellar, Thompson, & Rutkowski, 2017).

One of the concerns with global metrics and minimum benchmarks is how results will impact poor-performing countries (UIS, 2018h). League tables often lead to bifurcated reactions, heaping praise on high-performing countries while casting a dark shadow on low-achieving ones (Lockheed & Wagemaker, 2013). Countries near the bottom of an international ranking have reacted in various ways, including ceasing their participation in the assessment (Winthrop & Anderson Simons, 2013); withholding certain, less favourable results (Sellar et al., 2017); narrowing curricular policy to focus on tested subjects and question types (Heyneman and Lee, 2014); and attempting to 'emulate' practices believed to be prevalent in high-performing countries (Volante, 2015). The dangers of using a single measure to initiate or reform education policies have actually been highlighted by some of the biggest proponents of GLMs. The World Bank, for example, has cautioned that 'when a single metric becomes the sole basis for big policy triggers, the corresponding stakes may become dangerously high' (World Bank, 2017b, p. 93).

## 6 Assessments as Ready-Made Solutions

It is thus not a coincidence that as learning assessments expand at an unprecedented pace worldwide, rhetoric referring to the 'global learning crisis' reaches

new heights. Assessments not only provide a concrete gauge of the extent of the crisis, they also suggest solutions. For some, there is no difference between measuring the problem and solving the problem. As Pettersson and colleagues note, numbers on education can 'be transformed from representations of education into education per se' (Pettersson et al., 2016, p. 177). Test scores can be considered rationalised myths (Booher-Jennings, 2005). Abiding by an institutionalised, impersonal, and rationalised myth legitimates the behaviour of actors (Meyer & Rowan, 1977). In this sense, 'tests become a virtue *in and of themselves*' (Akkari, 2018, p. 10); all actors in education are expected to adapt their behaviour in response to, or for the sake of, improving test results.

Large-scale assessments both prompt the search for ready-made solutions and act as ready-made solutions themselves. Assessments can easily transform into simplified, prepackaged reforms that speed up the process of policy adoption and implementation (Lewis & Hogan, 2016). Access to ready-made solutions helps reassure the public that pressing learning problems are being addressed and resolved. The choice of a solution frequently aligns with already established cultural beliefs or views on education, often complimenting an ongoing intervention (Rosen, 2009).

In Chapter 12 of this volume, Clara Fontdevila recognises that assessments act as 'the policy *solution* to an institutionalized *problem* ... the learning crisis'. This is in part due to the perceived alignment between assessment and common values in education that prioritise some types of knowledge over others and a belief that the process of participating in a robust standardised assessment denotes a modern education system that is reflective, willing to learn, and based in science. Assessments capitalise on the prioritised position of academic intelligence. Subjects understood to require metacognitive skills are given more weight in assessments, while supposedly less cognitively demanding subjects like visual arts, ecology, or social studies are minimised (Baker, 2014). This is evident in the rise of sciences and mathematics as subjects for all students at the end of the 20th century (Kamens & Benavot, 1991; Kamens, Meyer, & Benavot, 1996).

Additionally, the application of assessments speaks to the widespread faith in science. The view of education 'as a "technical" science that can be studied, rationalized, and quantified' (Wiseman, 2010, p. 18) makes it difficult for policymakers to question the scientific results emanating from assessments (Rosen, 2009). Quantitative results are considered more accurate and trustworthy than summative pronouncements on the state of education or 'subjective' evaluations by teacher associations or school leaders (Wiseman, 2010). Cross-national assessments, which are considered a technically robust, valid scientific measure of academic knowledge, are perfectly positioned for countries

seeking to gain international legitimacy and demonstrate their commitment to quality education.

## 7 International Organisations and the Post-2015 Education Agenda

The view that internationally comparable learning assessments are the necessary means to both identify and remedy the learning crisis is bolstered by international agencies, regional organisations, and many NGOs (Boli & Thomas, 1999; see also Chapter 12 by Clara Fontdevila). Organisations such as the IEA, World Bank, OECD, UNESCO, and UNICEF have historically competed for supremacy in the global policy arena (Mundy, 1999). In the post-2015 landscape, the World Bank and the OECD have used their perceived technical expertise to solidify their position.

The World Bank's Education Sector Strategy 2020, *Learning for All* (World Bank, 2012b), illustrates the dominance of outcome over input in the organisation's thinking. In comparing this strategy, released in 2011, with the prior strategy from 1999, Joshi and Smith (2012) find a near 100% increase in terms associated with a testing culture. In practice, the World Bank spent the time between the two strategies adding assessment programmes to their funding packages. Between 1998 and 2009, the World Bank funded 166 projects with an assessment focus across 90 countries. In a quarter of the countries, the Bank funded participation in an international assessment. By organising the 166 projects into three-year periods equated with PISA assessment cycles, participation in international assessments grew from 7.1% (1998–2000) to 27.3% (2004–2006). See Figure 11.1 (World Bank, 2012a).

World Bank support for learning assessments can also be seen in the Systems Approach for Better Education Results (SABER) tool. Developed in 2011, SABER is designed to capture comparable information on education policies to drive institutional change through the creation of a national education report card. As part of their overall grade, countries are judged against their participation and implementation of learning assessments. Those that do not participate in international assessments receive lower marks (Bruns, Filmer, & Patrinos, 2011). Although voluntary in nature, the public shaming and increasing link between participation and funding opportunities suggests SABER acts as a normative guide to the 'right' kind of reform (Smith, 2014).

Similarly, the OECD has created a platform in which statistics are viewed as objective and, therefore, the results of their assessments (PISA and the Programme for the International Assessment of Adult Competencies [PIAAC]) produce indisputable scientific evidence (Martens, 2007). Through PISA, the

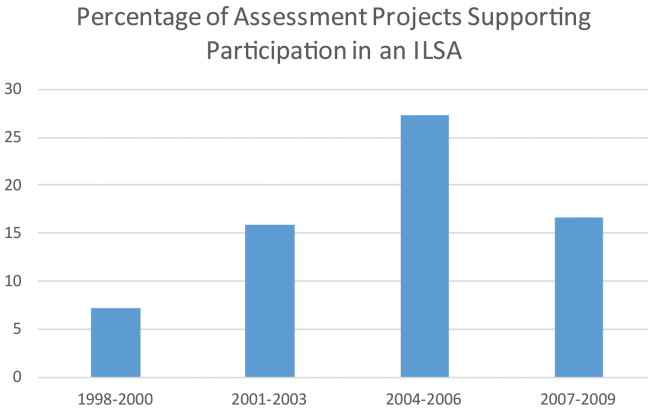


FIGURE 11.1 Changing inclusion of International Large-Scale Assessments (ILSAs) in World Bank projects, 1998–2009 (Source: World Bank, 2012a)

OECD promoted a performance culture (Bieber & Martens, 2011), which redefines competence and mastery in education. In the OECD's long-term strategy for PISA, 'the yardstick for educational success is no longer simply improvement against national standards, but against the best-performing education systems worldwide' (OECD, 2013, p. 4). In the run-up to SDG 4, OECD representatives sought ways to promote PISA as a global learning metric and expand its reach worldwide. To foster the participation of developing countries, the OECD piloted PISA for Development (PISA-D), which was designed to promote more robust national assessments through institutional capacity building. As one of PISA's six long-term objectives, efforts are underway to align national assessments with PISA scales. This has included workshops at the national level to develop PISA-like assessment instruments. To date, evidence that national assessments have been adapted to follow PISA proficiency levels can be found in China, Norway, Canada, Ecuador, and Paraguay (Addey, 2017). Although PISA was not ultimately endorsed as the global metric for measuring SDG Target 4.1, the OECD aspires for universal participation in PISA by 2030 (Ward, 2016). Signing India to PISA in 2021 is a step in that direction. In the meantime, the OECD is tracking progress on most SDG 4 targets of its own member states by repositioning various OECD indicators, including PISA results, as measures of SDG 4 targets (OECD, 2017a).

### 7.1 *Assessment Funded as Quality in Development Projects*

As noted above, assessment results are increasingly used as a proxy for quality in aid projects (Kamens & McNeely, 2010), reinforcing the testing culture and increasing low- and middle-income country participation in assessments

(Akkari, 2018). Donors are making funding decisions based on whether measurable impact can be identified (Winthrop and Anderson Simons, 2013). All too often, education is compared with the health sector, with donors demanding an easily implemented and communicated indicator (Hanushek & Edwards, 2017). However, unlike having children vaccinated or using bed nets to reduce the spread of malaria, education interventions are not easily packaged and implemented, nor can they be universally applied in every context (Snilstveit et al., 2016). Pressure to comply with the 'common sense' belief in assessment can be high. For example, in 2011 a major donor of the Global Campaign for Education (GCE) pulled its financial support when GCE failed to support a single, early-grade reading metric and instead continued its focus on rights-based education (Edwards, 2016).

Aid agencies are playing an influential role in the creation and legitimization of global learning metrics. The UK Department for International Development's (DFID's) funding for the 'Best Education Statistics for Improved Learning (BEST)' is an illustrative case. Running from 2013 to the end of 2017, the £6.4 million programme supported UIS's production of global learning outcomes, the OECD's PISA for Development, and the Global Education Monitoring (GEM) Report with an aim 'to help ensure education policies and programmes that are evidence based, with a greater focus on learning and ultimately contribute to disadvantaged girls and boys achieving improved learning outcomes' (UK DFID, 2016). In its 2016 review, DFID commended UIS for 'publishing data on learning outcomes for the first time', and with the establishment of the Global Alliance to Monitor Learning (GAML) and concept papers for universal learning scales, concluded that UIS 'exceeded expectations for work towards a global learning metric', although the planned work on equity was 'slightly behind schedule' (UK DFID, 2016). DFID recommended that it 'should track and support UIS's work on taking forward the Education Financing Commission's recommendation to develop a global lead indicator for education' (UK DFID, 2016).

To further intensify the link between assessment results and funding, some aid agencies have started to implement results-based financing in education. Many of the largest global donors, including the Asian Development Bank, the European Commission, Global Partnership for Education (GPE), DFID, and the World Bank, support some form of results-based financing, in which funding is released in proportion to the level of results attained (UNESCO, 2017b). Results-based financing is meant to spur progress by rewarding improvements in learning (Savedoff, 2016). Results-based financing using test scores typically provides funds to the national government, although this is not always the case. In Bangladesh and Chile, private providers were allocated funding for students who passed tests (Savedoff, 2016). This represents an increasing trend

toward local performance-based pay, shifting resources away from government provision (Savedoff, 2016).

The World Bank formally adopted a Program for Results (PforR) instrument in 2012 (Savedoff, 2016), after which it committed to doubling its results-based financing in the education sector to US\$5 billion between 2015 and 2020 (World Bank, 2015b). As of April 2016, the World Bank had initiated 37 PforR programmes, only two of which were in education. The largest education programme – the multi-donor ‘Big Results Now in Education’ programme in Tanzania – included a dedicated US\$122 million through the World Bank’s PforR instrument. In the Tanzania programme, six disbursement-linked indicators trigger the release of payment. The sixth indicator is improvement in student learning as demonstrated through a reading assessment. If this indicator were fully met, US\$16 million, or 13% of the total World Bank loan, would be released (Savedoff, 2016).

DFID piloted payment for results programmes in Ethiopia and Rwanda, which focussed on taking and passing competency tests (Savedoff, 2016). The programme in Ethiopia, to which DFID committed £30 million, continued a tradition in other sectors of linking British aid to results (Perakis & Savedoff, 2015). Launched in 2012 it rewarded the government for increasing the number of students sitting for the lower secondary exam and the number of students passing. Greater amounts were paid for girls sitting and passing the exam and for students from poorer regions. Tying payments to results at the end of lower secondary education was based on the assumption that sitting for the exam would be ‘a good proxy for completion, and resulting test scores provide information about the quality of the education’ (Perakis & Savedoff, 2015, p. 28).

The fund was expected to release £10 million per year for three years (Savedoff, 2016). In the first year £900,000 was released, in the second year £5.6 million, and in the third year £9 million. Poor performance in the first year was a challenge for DFID as the Ethiopian government treated the unreleased funds as lost revenue and a form of punishment (Perakis & Savedoff, 2015). Following the project’s conclusion, a DFID-commissioned evaluation found that the observed outcomes could not be attributed to the programme, which was not continued (Perakis & Savedoff, 2015).

## 8 SDG 4: The Global Goal on Education

### 8.1 *Creating and Then Undercutting an Ambitious Global Agenda in Education*

Since the SDGs came into force in January 2016, there appears to be a growing disjuncture between the espoused broad principles of the SDG agenda – namely,



equity, inclusion, and quality, which countries and institutions of widely differing ideological inclinations agreed to – and how these principles compete for supremacy during local policy creation and resource allocation (Clark et al., 1999; Smith, 2018). For example, SDG 4 calls on schools to develop multi-dimensional learners with wide-ranging knowledge and skills, but in practice, the content of education that countries are asked to assess is narrowly defined. The logic behind the process of delimiting the purposes of education appears reasonably sound: one cannot focus on all things at all times, and it makes sense to allow political and practical considerations to determine which principles gain traction and which do not.

And yet, there appears to be more than political expediency and pragmatism at work. In the run up to the new 2030 Agenda, representatives of particular education stakeholders were encouraged, sometimes forcefully, to jettison contentious proposals and strive toward consensus. During the SDG negotiations, different groups from the humanistic camp proposed targets and indicators that were later dropped or altered. In 2014, the Muscat Agreement proposed goals for both education financing and provision of trained teachers (UNESCO, 2014a). The draft Goal 7, requiring all countries to dedicate 4–6% of their GDP or 15–20% of their public expenditure to education, was subsequently discarded. Draft Goal 6, ensuring all learners are taught by ‘qualified, professionally-trained, motivated and well-supported teachers’, was eventually downgraded to a means of implementation and reworded to ‘substantially increase the supply of qualified teachers’. Supporters of adult education were told to get under the big tent idea of ‘lifelong learning’ and were left without any explicit mention of adult education in the targets (Benavot, 2018a).

Although it may not have been apparent at the time, the technical and political processes of operationalising targets and identifying indicators began to undercut the collective vision of an ambitious, holistic global agenda in education (McGrath & Nolan, 2016). Extensive consultation and consensus-building gave way to specialised experts leading technical discussions in which powerful and vocal international actors became more actively involved (Unterhalter, 2019). In meeting the ‘requirement’ for internationally comparable, quantifiable indicators (King, 2017), most of the proposed indicators in education were weakly aligned with the intended scope of the targets (Johnston, 2016). Among the 43 SDG 4 indicators that were proposed, only 11 came to be defined as global indicators with important measurement and reporting implications. Monitoring progress on SDG 4, as reflected in the UN’s annual SDG report, was limited to data on the 11 global indicators (King, 2017). Countries were not required to collect or report data on the 32 thematic indicators.<sup>3</sup> This suggests that, in practice, more countries are likely to pay closer attention and allocate more

resources to the monitoring and reporting of the global indicators (Smith, 2019), even though the UN Secretary General recognised that such indicators are ‘unlikely to fully satisfy the needs of communities’ (UIS, 2018h, p. 11).

## 8.2 *Capturing Quality in the SDG 4 Agenda*

As the results of international assessments are increasingly valued as objective measures of education quality, it should come as no surprise that the principle of quality is narrowly tied to assessments results. A pronounced utilitarian turn has taken root. As a result, discussions of SDG 4 targets, once they are operationalised and implemented, give minimal attention to equity and inclusion, and place the latter two in competition for scarce sources. After securing desirable formulations of SDG targets, the more humanistic approach voiced by civil society is being marginalised in practice (Doble, 2015). The instrumental view that strong test results, narrowly construed as proficiency in reading and mathematics, promote economic competitiveness continues to gather steam, thereby undermining the original intent of SDG 4 and the role of education as a driver of progress in other SDGs (Brissett & Mitter, 2017; King, 2017; Unterhalter, 2019; see also Chapter 9 by Yusuf Sayed and Kate Moriarty).

It is true that Target 4.7 captures a more humanistic, rights-based understanding of quality. However, that target includes multiple, often contested themes like global citizenship, sustainable development, human rights, gender equality, and cultural diversity, each of which embodies unique measurement challenges. For these and other reasons, the importance and value of Target 4.7 relative to other targets are being undermined (Brissett & Mitter, 2017). Unlike Target 4.1, which benefited from well-established measurement instruments, at least for the end of lower secondary education, approaches to measuring Target 4.7 themes have been few, uneven, and lacking consistent definitions (Unterhalter, 2019; UNESCO, 2016b).

Central to the remaking of the quality principle in SDG 4 is the role played by UIS. Due to several factors, including funding priorities and the divide between global and thematic indicators, UIS has focussed squarely on measuring the global indicator for Target 4.1 (Smith, 2019). Armed with a mandate to provide comparable data on SDG 4 indicators and establish robust methodologies for measuring each indicator, UIS has been the key player in creating consensus on learning metrics (see Chapter 12 by Clara Fontdevila in this book). Pushing to expand country coverage on SDG Indicator 4.1.1, UIS estimated that by 2017 only one third of countries had participated in a cross-national assessment of sufficient quality to allow for reporting (UIS, 2017b). To support and provide

legitimacy for the measurement of SDG 4 targets and indicators, UIS led the establishment of several expert groups.

The GAML constitutes the largest and most active group. It focusses on SDG 4 targets with learning outcomes – namely, targets 4.1, 4.2, 4.4, 4.6, and 4.7 (UIS, 2016b) – and includes many of the same actors that participated in the Learning Metrics Task Force. As membership in the GAML is based on self-funded participation, donors dominate meetings and have played a significant role in directing the focus toward Target 4.1. At their May 2017 meeting, the GAML concluded that a global reporting scale should be developed by mapping national, regional, and international assessments to a common metric (UIS, 2017b). To take advantage of the wave of international assessments planned for 2018 and 2019, UIS and the GAML made the production of the necessary methodology, global reporting scale, and metadata an urgent priority to be met by December 2018 (UIS, 2017b).

Contrasting the activity of the GAML with the Inter-Agency Group on Educational Inequality Indicators (IAG-EII; originally the Inter-Agency Group on Disaggregated Education Indicators) helps highlight the prioritisation of learning assessments. The IAG-EII was developed ‘in response to the call for a greater focus on equity in the global post-2015 education agenda and for more efficient use of available information’ (UIS, 2016a, p. 3) and counted UIS, UNICEF, and the World Bank as its founding members. Compared to the GAML, the members of the IAG-EII appear to be taking a more relaxed approach to meeting their relatively fewer goals. The concept note for the group states that over a three- to five-year period they seek to harmonise the definition of individual characteristics such as sex, location, and wealth. The group aims ‘to summarize periodically (if possible, annually) the main findings on the key indicators in a report’ (UIS, 2016a, p. 4). However, the first report, originally targeting a December 2017 publication date, has yet to be published on the group’s website.

In comparison, the GAML Results Framework outlines ambitious production with 31 separate outputs planned around assessment between February 2017 and the end of 2018 (UIS, 2017f). One could argue that it makes little sense to compare these two entities: GAML is larger and includes more self-funded members. That, however, is precisely the point. The energy and resources connected to the GAML demonstrate the prioritisation of learning assessments – assumed to be a measure of quality – over disaggregated equity indicators. Finally, one may argue that Target 4.1 is not all that the GAML covers. Yet, in evaluating the GAML’s Results Framework, the combined targets directly tied to 4.2, 4.4, 4.6, and 4.7 total just less than one-third of the more general outputs on learning assessment (10 relative to 31).

## 9 What Does This Mean for Equity and a GLM?

The enormous time and effort expended to operationalise SDG 4, primarily by assessing reading and mathematics proficiency levels on a GLM, means that many quality and equity issues have taken a back seat. Improving equity in tested reading and mathematics levels does little to improve broader quality and equity concerns in education. As Slee (2013, p. 6) pointed out, more humanist values of education, such as inclusion and equity, are often ‘inaudible when located amid more strident educational discourses’, such as standards and performance rankings. One obvious example of this is the inattention to stalled out-of-school numbers – the premier access measure under EFA (Hanushek & Edwards, 2017). As learning outcomes are prioritised, some argue that ‘enrollment is no longer the main issue’ (Savedoff, 2016); instead countries are expected to focus on the low level of learning taking place in schools. Learning is narrowly understood in this context as school-based knowledge captured through comparative learning assessments. According to UIS and others, ‘learning goals and targets in the post-2015 agenda *will only be meaningful* if they are underpinned by empirically derived common numerical scales that accommodate results from a range of different assessments of learning outcomes’ (UIS et al., 2014, p. 1, emphasis added). This helps explain the concentrated, almost relentless efforts to develop GLMs that measure the global indicator for Target 4.1, while measurement strategies for other SDG 4 indicators – even those connected to learning – languish by comparison.

While quantitative experts continue to tinker with alternative methods to map results from different learning assessments onto a single global metric, we raise a final issue: what might be the impact or possible unintended consequences of GLMs on the global education agenda, especially in relation to equity and quality. We begin by noting that in 2017, for the first time, UIS estimated that over 617 million children and adolescents around the world were not meeting minimum proficiency levels in reading and mathematics (UIS, 2017i). Supported by DFID and the work of the GAML, this top-line number is, in part, UIS’s response to a recommendation from the International Commission on Financing Global Education (Education Commission, 2016)<sup>4</sup> and a reflection of UIS’s desire to meet donor demands and drive advocacy. This figure seeks to provide a snapshot of the percentage of children and youth who have not acquired basic foundational skills, worldwide and by geographic region. UIS further drew attention to responsibility at the local level. In identifying three potential causes for the ‘fact’ that globally 58% and 56% of the relevant age group will not reach minimum proficiency levels in reading and

mathematics, respectively, blame was squarely placed on schools and in classrooms (Montoya, 2018).

The World Bank introduced an updated global dataset on education quality in January 2018. The full database contains data from 163 countries spanning the years 1965–2015. It illustrates one potential iteration of a global learning metric by calibrating international and regional assessments to an overall metric through doubleton countries – countries that participated in both an international and a regional assessment. Over time, the United States is used as an anchor country, as it is the only country that has participated in the full set of assessments since 1964. Results can be disaggregated by gender, location, immigrant status, and home language (Altinok, Angrist, & Patrinos, 2018). Harmonised test scores from the global dataset on education quality constitute one of five components used by the World Bank to calculate their newly launched Human Capital Index (HCI) (World Bank, 2018a).

The abovementioned examples demonstrate that as figures derived from GLMs take centre stage, the needs of marginalised groups of children are likely to go unnoticed and underappreciated. Consider out-of-school children, who are no longer highlighted as one of the most pressing policy concerns. In 2016, 263 million school-age children remained out of school (UIS, 2018d). At the primary level, about 40% of the 61 million out-of-school children are never expected to enter school (UNESCO, 2016b). Furthermore, providing access for the most marginalised is more expensive; without consistent focus and funding, these children are likely to remain excluded (Smith, 2019).

Most large-scale assessments are not designed to address the learning challenges faced by out-of-school children or youth. Nearly all international and regional assessments omit out-of-school children from their sampling frames (Winthrop & Anderson Simons, 2013). As Sellar et al. (2017) suggest, not only does this disadvantage those out of school, but it can also lead to a misinterpretation of results. For example, PISA tests 15-year-olds with results usually generalised to the entire age cohort. However, in their examination of 16 countries that participated in the 2012 PISA, out-of-school rates meant that one in five 15-year-olds were not included in the sampling frame. In Costa Rica, nearly 50% of 15-year-olds were not in school, and thus their learning levels were not captured by the PISA assessment.

The learning measure included in the aforementioned HCI is based on results from regional and international assessments without concern for those out of school. This means that in national contexts where out-of-school levels are relatively high, the index overestimates a major component. Some approaches to constructing GLMs seek to overcome this issue by mapping household surveys and citizen-led assessments to the global scale (Montoya &

Hastedt, 2017). However, unless all countries take part in surveys that capture children both in and out of school, true measures of out-of-school children, including their learning levels, will be biased. In the end, GLMs simply do not provide the country-specific information needed by policymakers to understand the nature of the learning challenges facing marginalised children and thus identify potentially effective policies.

Interestingly, in its calculation of the global number of children and youth who will not meet minimum proficiency levels, UIS includes estimates of those out of school (UIS, 2017i). However, the presentation of the global figures almost entirely focusses on learning deficits and not on identifying effective policies to ensure access and completion of the most vulnerable and marginalised populations (Smith, 2019). One report estimates that the world's out-of-school population constitutes about 15% of the total number of children lacking sufficient literacy and numeracy skills (UIS, 2017i). Breaking this global number down by region, a recent UIS blog illustrates that among children not meeting minimum proficiency, the percentage that are out-of-school is lower in Asia, Latin America, and Sub-Saharan Africa than in Europe and North America (Montoya, 2018). In looking at how these numbers are represented, there are concerns that to address the 'learning crisis' policymakers may focus attention on the group representing the largest percentage of children and youth contributing to the 'crisis', such as those in school in Sub-Saharan Africa. This would diminish attention on the still severe out-of-school issue, especially in regions where the challenge is most pressing.

Another critical implication stems from the fact that current learning assessments capture a limited range of learning domains. Education has always valued the knowledge of some learners over others. Assessments tend to privilege academic knowledge and cognitive skills in a few subject domains: language, mathematics, social and natural sciences (Benavot, 2018b). These tendencies undergird 'a meritocratic ideology' that 'has not only brought about assessment practices that enable and promote some, but not other, educational activities' but also 'sustains and legitimizes educational distribution of life chances for different individuals' (Pettersson et al., 2016, p. 197). The cultural knowledge and competencies of learners from ethnic, religious, or linguistic minorities, who are frequently among those excluded from school or unable to complete a full cycle, are often unrecognised, untested, or both. There is little information in current GLMs that would provide policymakers and educators with information about the challenges faced by such learners.

GLMs favour knowledge conveyed in school during the initial phases of life-long learning. The diverse types of knowledge and skills, which are learned or reinforced over the life course both in and out of school, are not assessed

(Benavot, 2018b). For example, the World Bank's dataset relies on regional and international assessments of school-based knowledge. While the UIS top-line number includes out-of-school children, it makes a false and highly problematic assumption that no learning takes place outside the confines of school. All out-of-school children are assumed, by definition, to be lacking minimum proficiencies in foundational skills.

This discussion raises a deep concern as to the purposes of assessment: what is being prioritised in assessment results and who benefits from such assessments? Countries have defined different educational purposes and curricular priorities, which should be taken into account in assessment frameworks (Winthrop and Anderson Simons, 2013). Not all assessments can and should be used for the same purpose (Benavot & Köseleci, 2015). And yet, among large scale assessments that benchmark proficiency levels, UNESCO's Technical Cooperation Group (UIS, 2018h) suggests there are two main purposes: improved learning and hierarchical control. Some have debated these purposes of GLMs, questioning the utility of a simple tool or benchmark for improving student performance or whether they serve as instruments to punish countries through rankings (Winn & Goebel, 2017). Eric Hanushek suggests that GLMs help us 'think about judging the education' in different countries (Cavanagh, 2018); the World Bank supports GLMs as a form of cross-country comparison that helps 'generate accountability for learning' (World Bank, 2017b, p. 97).

Well-designed learning assessments should guide policy interventions to improve quality and student learning. However, while global rankings may drive policy or identify 'stages at which policy interventions may be required', they provide preciously little information to guide potential interventions or improve the quality of instruction (UIS et al., 2014, p. 12). This is especially true in those countries near the bottom of the league tables that tend to focus 'only on their rankings rather than on using the results to stimulate reflection on how they might improve their system' (Lockheed & Wagemaker, 2013, p. 297). Even for willing systems, the decontextualised piecemeal information contained in rankings is insufficient for effective action. Of the results available, they are not provided in a timely manner to stakeholders who can make use of them (World Bank, 2017b). As Edwards suggests, if 'it was about improvement then the information would feed its way back into the hands of the people that are best positioned to make decisions to improve' (Hanushek & Edwards, 2017).

Adapting international learning assessments to the national context potentially enhances their ability to address learning needs. 'Not aligning metrics to national policy and curricula will reduce their use and usefulness in informing



policy development and supporting classroom interventions as they diverge from countries' needs and priorities' (UIS, 2018h, p. 45). National and, to a lesser extent, regional assessments are more likely than international assessments to capture context-specific factors that foster or impede learning. National assessments are less expensive to administer, less likely to lead to shallow calls for wholesale reform, and more likely to capture the implemented curriculum in a country and its role in student learning (Kamens & Benavot, 2011). Unfortunately, the current vision for GLMs relies, almost exclusively, on international and regional assessments. If a guiding principle in the push for GLMs is to establish no new tests (UIS, 2017b), then it remains unclear if developing countries will have the expertise, funds, and capacities needed to strengthen national assessments to acceptable quality.

If the prevailing purposes of assessments are neither feasible nor desirable, why do developing countries participate? For many, participation acts as a visual marker of a serious commitment to improving the education system. For example, the post-Pinochet government in Chile decided to participate in international assessments, thus demonstrating a functioning government to its citizens (Kijima & Leer, 2016). For other countries, external pressure to be included in the global wave of assessments is strong. As one official in Paraguay stated when asked about his country's participation in PISA-D, 'Not being on the information map in the 21st century is unbearable' (Addey & Sellar, 2018). Developing countries, as latecomers to international assessments, tend to participate to establish legitimacy by 'doing what is expected of them by their individual and institutional peers' (Wiseman, 2010, p. 2). And, as previously mentioned, participation is increasingly an integral component of aid packages. The promise that country participation in GLMs will be voluntary (Motivans, 2014) rings hollow, given these many pressures on developing countries.

## 10 Conclusion

This chapter focussed on the construction and legitimization of global learning metrics in the context of SDG 4. It argued that learning assessments are being promoted as a ready-made solution with a viable methodology to measure a simplified notion of quality – namely, whether students achieve minimum proficiency levels in reading and mathematics in primary and secondary education. This abiding faith in measurement as a pathway to solving the 'global learning crisis' is rooted in a culture in which tests are seen as objective measures of individual and national effort, and higher test scores are associated with increased economic prosperity.

Currently, the global scales for monitoring proficiency levels have yet to be finalised. In the interim, early versions of GLMS raise concerns as to their impact on equity issues: Will the obsession over measuring and monitoring learning diminish the importance of, or effectively marginalise, equity concerns in SDG 4? Will GLMS undermine support and funding for the 263 million children and youth who are excluded from school? Will country efforts to improve foundational skills in literacy and numeracy undercut innovative policies and practices to ensure that all out-of-school children and youth gain access to a full and rich education and enjoy its benefits?

The dramatic depiction of a world in which hundreds of millions of children and youth are not learning, and the implications of this for future prosperity, downplays whether countries are taking concrete steps to universalise completion of primary and secondary education and ensure that 12 years of schooling are free (specifically, fee-free). Paradoxically, the global figure of more than 600 million not acquiring foundational skills is itself based on out-of-school figures for children and youth who are *assumed* to not be learning. This suggests that the 'learning crisis' in regions like Sub-Saharan Africa is driven in no small measure by those not attending school and those not completing a full cycle of schooling. Similarly, the World Bank's global dataset on education quality draws from international and regional surveys on children in school, altogether omitting those out of school.

This chapter also raises concerns about the relevance and utility of GLMS to countries and education professionals. Countries have been the movers and shapers of the 2030 Agenda for Sustainable Development. Country representatives led discussions over the formulation and scope of the SDGs, and they committed their governments to implementing this ambitious vision and comprehensive set of targets. The policies and actions they have and will put in place, and the partnerships they empower, will determine, in the final analysis, how much progress will be realised in the coming years. Of concern is whether GLMS meet the specific needs of countries: do they provide useful, contextualised information for feasible reform efforts? And when countries find themselves at the bottom or near the bottom of the new league tables, what steps will they take? Narrow the curriculum or teacher preparation? Exclude poor achievers from the assessments? Weaken the more challenging multilingual approaches to teaching and learning?

The intense focus on outcomes, especially learning outcomes, in the broad SDG 4 agenda is quite clear. Many of the 10 SDG 4 targets and the 11 global indicators emphasise important outcomes of schooling – specifically, foundational skills, employability skills, youth and adult literacy, knowledge and skills for global citizenship and sustainability, and more. These results-oriented

indicators serve as an accountability framework for determining country progress. Though a broad range of actors contributed to the formulation of goals and targets, in developing indicators, it has been experts, international agencies, and donors, all legitimate actors, whose voices are the strongest. The views of country leaders, civil society representatives, and educators, who typically support a more humanistic understanding of education, have taken a back seat. The dialogue and decisions around putting SDG 4 into action seems to be dominated by those supporting an economic-oriented, utilitarian role for education. Given the time, collective effort, and funding needed to consider and pilot alternative measurement strategies, the longer-term impact of these actors and their actions should not be minimised. Consciously or not, their decisions as to which concrete procedures should be employed to reflect education realities on the ground have effectively prioritised which targets and indicators will gain visibility and which will not.

As the SDG era progresses, the issues outlined in this chapter are unlikely to be attenuated. The tensions between different camps and viewpoints, especially at the country and regional levels, are likely to become more palpable. Simplifying quality education to the lowest common denominator – namely, minimum proficiencies in foundational skills – and using this measure to dominate the policy discourse and donor priorities, GLMS threaten to broaden existing inequalities, valuing some forms of knowledge over others, and benefiting those already invested in large-scale assessments.

## Notes

- 1 These proficiency levels are reported for at least one data point since 2012 (see <http://data.uis.unesco.org/>). To access these data, look under 'SDG4', 'Target 4.1', and 'Indicator 4.1.1', entitled 'Achieving at least minimum proficiency level in reading at the end of lower secondary education'.
- 2 Three approaches are currently being considered: statistical recalibration, social moderation, and the Rosetta stone approach (Gustafsson, 2018).
- 3 Strictly speaking, reporting on both global and thematic indicators is voluntary, although more attention is paid to the former than the latter.
- 4 From Recommendation 1: 'To galvanize attention globally, a single global indicator of learning should be agreed on to complement national measures of learning. The international community should track, rank, and publicize countries' progress in getting all children learning' (Education Commission, 2016, p. 17).